

Philosophie de l'esprit et neuropsychologie

in *Les Méthodes de la neuropsychologie*, ed. F. Eustache, B. Lechevallier, et F. Viader, de

Boeck, Bruxelles, 2000, pp. 39-53

Pascal Engel

Il y a, de prime abord, trois manières de concevoir ce que l'on appelle aujourd'hui la "philosophie de l'esprit" dans sa relation aux sciences cognitives. La première, traditionnelle, considère que les problèmes portant sur la nature des phénomènes mentaux sont essentiellement des problèmes métaphysiques, peut-être comme tels insolubles, et qui relèvent en tous cas d'analyses conceptuelles purement a priori, qu'aucune enquête empirique ne peut ébranler. La seconde, non moins traditionnelle mais opposée à la précédente, consiste à prendre pour argent comptant les avancées des sciences cognitives dans l'étude de ce qu'on appelle "l'esprit", et à traiter la philosophie comme une sorte de métadiscipline qui examine les méthodes et les concepts de ces sciences, en montrant comment les questions métaphysiques classiques s'y trouvent résorbées. La troisième, plus subtile, prend acte des progrès scientifiques en ces domaines, mais assigne au philosophe le rôle de dénoncer les mythologies et les confusions conceptuelles que les sciences cognitives, comme les autres sciences, ne peuvent manquer d'entretenir¹. À mes yeux, aucune de ces manières de concevoir la tâche d'une philosophie de l'esprit n'est la bonne. La philosophie n'est ni la servante des sciences (cognitives ou autres) qui ramasserait les détritits conceptuels tombés sous

¹ Si l'on veut fixer les idées, disons qu'un auteur comme Mc Ginn (cf. par exemple son livre *the Character of Mind*, Oxford University Press, 1983) est un tenant de la première position, des auteurs comme les Churchland (P.S. Churchland, *Neurophilosophy*, MIT Press, 1986, tr.fr. *Neurophilosophie*, Paris PUF 1999, P.M. Churchland, *Matter and Consciousness*, MIT Press 1991) sont des tenants de la seconde position, et des auteurs wittgensteiniens, comme P.Hacker et G. Baker (ex. leur livre *Language, Sense and Non sense*, Oxford, Blackwell, 1985) ou Vincent Descombes (1995) sont des tenants de la troisième attitude.

la table de du festin scientifique, ni n'a de titres à prendre une attitude grand seigneur ou de redresseuse de torts. Elle a ses problèmes propres, mais elle ne peut, le plus souvent, les traiter que par une interaction avec les sciences. Je voudrais essayer d'illustrer la manière dont ce genre d'interactions peuvent se produire à propos d'une discipline cognitive particulière, la neuropsychologie.

La neuropsychologie cognitive a spécifiquement pour objectif d'analyser la nature des processus cognitifs normaux à la lumière des désordres cognitifs. Elle se caractérise par un certain nombre d'hypothèses méthodologiques et théoriques, bien mises en valeur par le livre de Shallice, *Symptômes et modèles en neuropsychologie* (Shallice 1995). La première, que je viens de mentionner, est que l'analyse de déficits cognitifs et de diverses lésions doit se baser sur l'analyse des cas de cognition normale, mais qu'en retour celle-ci peut être éclairée par les cas pathologiques. En particulier, si l'on met en évidence que dans des cas de lésions certaines structures et fonctionnements cognitifs sont préservés, alors même qu'ils ne devraient, selon certains modèles de fonctionnement cognitif normal, ne pas être préservés, cette découverte tend à infirmer ces modèles. On s'aperçoit en effet souvent que là où certains processus subissent des désordres, d'autres continuent d'opérer normalement. Si, par exemple, on découvre d'une part des patients qui sont aptes à la lecture de mots familiers, y compris irréguliers, mais inaptes ou peu aptes à la lecture à haute voix de mots non familiers ou de non-mots, et d'autre part des patients qui parviennent à lire à haute voix des mots réguliers et des non-mots mais échouent à lire à haute voix des mots irréguliers, cette double dissociation suggère qu'il y a deux systèmes distincts pour la lecture à haute voix, c'est-à-dire deux chemins distincts qui mènent de la reconnaissance des signes orthographiques à la phonologie (Coltheart, 1985). Ces découvertes reposent sur, et confirment en partie, une hypothèse théorique majeure: la structure de l'esprit est modulaire. Il y a des composants distincts, spécifiques, qui peuvent faire l'objet de dysfonctionnements sélectifs. Les lésions subies par diverses composantes

laissent inchangées les opérations des modules non atteints. Une troisième hypothèse, cette fois essentiellement méthodologique, consiste à privilégier l'étude de cas unique; en ce sens la notion de symptôme, qui repose sur l'étude de groupes de patients, est problématique. On peut enfin mentionner une quatrième grande hypothèse, que l'on peut appeler celle de l'autonomie relative de la neuropsychologie cognitive par rapport à la neurophysiologie. Ici cependant les avis sont plus divergents. Certains auteurs défendent une position réductionniste, et pensent que la neuropsychologie doit se fonder sur la neurophysiologie seule. D'autres à l'inverse soutiennent une version que Shallice appelle "ultracognitiviste" consistant à prêcher l'indépendance complète des structures modulaires par rapport aux structures anatomiques et neurologiques. D'autres encore, comme Shallice, défendent une position intermédiaire, qui se veut située aussi bien "à l'écart du Scylla de la dépendance exclusive d'une approche réductionniste classique qui repose uniquement sur les études de groupes de patients qu'à l'écart du "Charybde de la neuropsychologie ultracognitiviste" (Shallice, 1995: 255).

À elles seules, ces hypothèses sont suffisantes pour éveiller l'attention du philosophe. Celui-ci peut se demander, par exemple, sur quelles bases repose l'hypothèse de modularité employée ici, et si elle est distincte ou non de celle que des philosophes comme Fodor (1983) proposent. Il peut aussi se demander si le cognitivisme modéré d'un Shallice est compatible avec ce que les philosophes appellent une thèse fonctionnaliste en philosophie de l'esprit, selon laquelle la nature des états mentaux consiste en leur rôle causal ou fonctionnel, indépendamment de leur "réalisation" dans des structures neurophysiologiques particulières ou avec une forme plus traditionnelle de matérialisme selon laquelle les états mentaux seraient identiques à des structures neuronales. Mais ce ne sont pas ces questions qui m'intéressent ici. Je voudrais m'intéresser à des positions radicales défendues par certains philosophes qui sont bien plus critiques à l'égard des thèses sous-jacentes à la neuropsychologie que ne peuvent l'être les philosophes fonctionnalistes ou matérialistes, parce que ces positions radicales

mettent en question l'idée même que la neuropsychologie pourrait avoir la moindre pertinence pour l'analyse du mental².

Considérons une sorte de cas typique des analyses neuropsychologiques, celui d'une prosopagnosie. De Haan et ses collaborateurs (1987 ; voir aussi Young et de Haan, 1988) décrivent le cas d'un patient, P.H., dont le déficit était très spécifique: il parvenait à reconnaître les visages de certaines personnes à partir de la présentation orale et visuelle de leurs noms, mais pas à partir de leurs visages. De Haan et ses collaborateurs proposent un modèle de la reconnaissance des visages selon lequel un visage est, à un premier niveau de traitement d'information, reconnu comme familier ou pas, et où une description est produite, puis comparée à une sorte de fichier de descriptions; ensuite où, à un second niveau, une information sémantique devient disponible, rapportant les descriptions à une personne ou à une de ses propriétés; et enfin où, à un troisième niveau, le visage est identifiée et reçoit un nom. Les expériences révèlent une structure de dissociation de ces trois niveaux chez P.H., c'est-à-dire la compatibilité entre un niveau où les descriptions sont atteintes bien que P.H. ne reconnaisse pas les visages; et l'existence d'un niveau de traitement de l'information sémantique lui même compatible avec la non reconnaissance complète des visages. Une expérience de catégorisation de noms permet d'établir ces résultats. Chez un sujet normal, on sait que la connaissance d'un visage familier interfère avec les tâches de catégorisation de noms. Par exemple, si on demande à un sujet normal de catégoriser le nom "Jacques Chirac" comme le nom d'un sportif plutôt que celui d'un politicien, ils présentent des temps de réaction plus rapides quand le nom est présenté seul que quand il est présenté avec le visage de quelqu'un appartenant à une catégorie différente.

De Haan et ses collaborateurs présentent à P.H. cinq conditions distinctes:

Même personne: visage et nom de la même personne.

² Je me suis beaucoup inspiré des discussions de Stone et Davies (1994).

Nom seul: seul le nom est donné.

Pas de rapport: le visage et le nom sont ceux de gens de diverses catégories, qui ne se ressemblent pas.

Relation de catégorie sémantique: le nom et le visage sont ceux de personnes distinctes qui ont la même occupation, mais qui ne se ressemblent pas.

Lien visuel: le nom et le visage appartiennent à des gens de différentes catégories, mais d'apparence physique semblable (ex. Jean Pierre Chevènement et Mr Bean).

Les réponses de P.H. à ces conditions sont en général plus lentes que pour le sujet normal. Mais pour lui comme pour le sujet normal, ses réponses dans les conditions Pas de rapport et Lien visuel sont plus lentes que dans la condition Nom seul. Pour lui, voir le visage de quelqu'un qui a une occupation différente inhibe la catégorisation du nom. P.H. du mal à reconnaître les visages, mais il a une reconnaissance tacite du fait que ce sont des gens qui ont une occupation distincte, car autrement on ne pourrait pas expliquer que le visage de Jean Pierre Chevènement ait inhibé la catégorisation de Mr Bean comme acteur.

D'autres travaux (Etcoff et al., 1992) montrent qu'un patient peut avoir une "connaissance tacite" (*covert knowledge*) des expressions faciales des gens dont il voit les visages, bien qu'il ne reconnaisse pas ces visages.

Selon certains philosophes, cependant, toute cette description, qui implique l'existence de processus mentaux particuliers, conscients ou non, est totalement erronée et illusoire, parce qu'il n'y a pas de processus mentaux constitutifs d'un état mental donné. Je peux, par exemple, essayer de me souvenir où se trouve mon cartable que j'ai perdu. Je peux passer en revue les lieux où je me suis trouvé avec lui, avoir diverses images, etc. Mais je peux parfaitement aussi essayer de me souvenir où j'ai laissé mon cartable sans que de tels processus et événements aient lieu. Ces événements ne définissent donc nullement mon activité d'essayer de me souvenir. L'idée n'est pas simplement qu'il n'y a pas de processus conscients de ce type, et qu'ils pourraient être inconscients, mais qu'il n'y a pas de processus du

tout, conscients ou non. C'est simplement une erreur de catégorie, ou une erreur conceptuelle, que d'associer un état mental quelconque avec un processus ou un événement qui se déroulerait dans la tête de quelqu'un.

On reconnaît ici une thèse qui a été défendue spécifiquement par Wittgenstein et un certain nombre d'auteurs à sa suite (en France par exemple, Descombes, 1995). Selon ces auteurs, les états mentaux se définissent essentiellement par les critères et les usages que nous en avons, dans la pratique la plus ordinaire. Ces critères sont fournis par leurs manifestations comportementales et par les dispositions physiques auxquelles ils font référence, c'est-à-dire par les moyens que nous avons de les vérifier. Cette thèse ne revient pas nécessairement à un behaviorisme. Certains concepts mentaux, comme ceux de douleur ou de sensations, peuvent renvoyer à des critères comportementaux quand ils sont attribués à autrui, mais aussi à des critères introspectifs quand nous nous les attribuons à nous mêmes. Un autre principe de la conception wittgensteinienne est que ces critères *définissent* les concepts mentaux en question, en sorte que si l'on venait à identifier ceux-ci à des entités ou à des événements particuliers, nous changerions tout simplement le sens des mots et nous parlerions d'autre chose. Ainsi le wittgensteinien Malcolm (1959) défend-il l'idée que même si la recherche neurophysiologique sur les rêves conduisait à identifier ceux-ci à des trames d'événements physiques dans le cerveau, on changerait le sens du mot "rêve" parce qu'on l'associerait à de nouveaux critères. On associerait un état mental dont le sens est intrinsèquement associé aux comptes rendus que nous donnons de nos rêves quand nous sommes éveillés, à des événements physiques: et en ce sens on ne parlerait plus de la même chose.

La question de principe qu'on est en droit de soulever face à ce genre de questions est la suivante: en quoi les critères que nous associons couramment à nos termes mentaux sont-ils réellement des définitions de ces termes, et pourquoi ces définitions ne pourraient-elles pas changer? Il n'y a certes rien, dans la conception wittgensteinienne, qui l'interdise, et rien qui s'oppose, par exemple, à ce que

l'humanité finisse par reconnaître comme correctes les positions de Damasio sur la nature des émotions, et cesse un jour de parler de "peur", d'"amour" ou de "joie", au bénéfice des descriptions neurophysiologiques de ces états. La thèse des wittgensteiniens n'est pas que cela ne pourrait pas se produire, mais que cela ne s'est pas produit. En ce sens, il n'y a pas nécessairement de désaccord avec une conception qui nous conduirait à associer au mental d'autres critères que ceux dont nous disposons. Le véritable désaccord est ailleurs. Il tient à deux thèses annexes des wittgensteiniens.

La première est qu'une description des phénomènes mentaux qui emploierait un vocabulaire neurophysiologique ou neuropsychologique, décrivant des événements par nature inconscients, cesserait par là même de parler de phénomènes mentaux. Elle parlerait, selon les termes de l'un des avocats les plus éloquents de cette théorie, des "*limites physiologiques du mental, mais pas des limites du mental*" (Descombes, 1995: 216). Pourquoi ne parlerait-elle pas de phénomènes mentaux ? Parce que, selon nos auteurs wittgensteiniens, les phénomènes mentaux se caractérisent par deux propriétés essentielles: ils sont *intentionnels*, c'est-à-dire portent sur des contenus sémantiques du type de ceux que nos rapportons habituellement au moyen d'attributions de ce que les philosophes appellent des attitudes propositionnelles, telles que des croyances, des désirs, ou des souhaits, et ils ont certaines caractéristiques *rationnelles*, au sens où ces contenus entrent dans certaines trames inférentielles permettant d'établir des relations entre ces contenus. Selon les wittgensteiniens, intentionnalité et rationalité vont de pair, car elles supposent que quelque chose n'est un phénomène mental authentique que si le sujet qui les a est capable de fournir les *raisons* de ses états mentaux particuliers. Il s'ensuit que toute description des états mentaux à un niveau qui serait *infra-rationnel*, ou *subintentionnel* ne parlerait plus d'états mentaux proprement dits, parce qu'il ne parlerait plus de raisons, mais de causes. Or, poursuit l'argument, s'il est légitime de penser que les causes d'un certain nombre d'états mentaux, tels que des perceptions par exemple, sont des *conditions* de ces états mentaux, il est illégitime de

supposer que ces conditions causales définissent les états en question. Ce qui les définit est leur relation aux raisons que les agents peuvent eux-mêmes donner. Le même argument permet d'écartier toute description du mental qui se réduirait à l'analyse d'un niveau de représentation mentale inconsciente qui ne pourrait pas être accessible à la conscience, et toute description du mental en termes strictement neurophysiologiques. Si l'on passait à ces niveaux de description, on changerait le sujet. Selon cet argument, parler de processus "cognitifs" qui existeraient à un niveau subintentionnel est une forme de contradiction dans les termes: car ou bien les processus en question sont réellement des inférences, et dans ce cas ils doivent illustrer l'activité de suivre des règles, qui ne peut être que consciente, ou bien ils sont descriptibles en termes neurophysiologiques, et en ce cas ils ne sont pas non plus des "inférences", mais de purs processus causaux, qui sortent par là même de l'"espace des raisons"³.

La seconde thèse des wittgensteiniens, moins explicite que la précédente, est que l'on ne peut déterminer la nature d'un phénomène mental, et par conséquent ses critères, que par rapport aux cas normaux de son application, c'est-à-dire en fonction de ce que la "psychologie du sens commun" ou "populaire" est en mesure de nous en dire. L'idée est ici que ce que c'est que croire quelque chose, désirer quelque chose, ou avoir peur, ne peut se comprendre que par rapport à des cas paradigmatiques ou typiques, dont la psychologie usuelle est le conservatoire.

Essayons de répondre à ces arguments à partir des cas neuropsychologiques mentionnés ci-dessus. Il ne s'agit pas de nier la différence entre la psychologie au niveau personnel et intentionnel et la psychologie au niveau subintentionnel, mais la thèse selon laquelle ces niveaux n'ont aucune interaction et que le second n'a aucune pertinence pour l'explication du premier est indéfendable. Dans le cas des tâches de catégorisation de noms, la psychologie populaire peut certes évoquer la notion de distraction: si un sujet normal reconnaît le visage présenté et sait que

³ Selon l'expression célèbre du philosophe Wilfrid Sellars (1963). Cette argumentation n'est pas nécessairement wittgensteinienne. Elle peut aussi s'autoriser d'une position "néo-cartésienne" du type de celle de Searle (1992), ou d'une argumentation davidsonienne (P. Engel 1996).

c'est celui d'un politicien, il n'est pas surprenant que cette connaissance interfère avec le fait qu'on le présente comme celui d'un acteur. P.H. partage avec le sujet normal l'effet d'interférence dans la condition *pas de relation*; mais la psychologie populaire ne peut pas expliquer le cas de P.H., car l'effet d'interférence intervient chez lui sans reconnaissance consciente du visage comme celui de Jean Pierre Chevènement, ni comme celui d'un politicien.

Mais ici nous n'avons pas deux explications qui n'ont aucun lien ni aucune pertinence l'une par rapport à l'autre. Car de deux choses l'une. Ou les explications en termes de nos concepts psychologiques de sens commun ne sont pas des explications causales — mais des explications par les raisons qui en sont foncièrement distinctes — et en ce cas le sens commun n'a tout simplement pas d'explications pour ces phénomènes. Ou bien ce sont bien des explications causales, ou qui des explications qui visent à l'être. Et en ce cas elles sont en compétition l'une avec l'autre. Et si elles le sont, l'explication en termes de traitement inconscient d'information est pertinente pour l'explication des effets d'interférence observés chez les sujets normaux⁴. Dans la première hypothèse, la conception wittgensteinienne du mental n'a tout simplement pas de place pour des explications de phénomènes mentaux pathologiques ou anormaux, ce qui n'a rien de surprenant au regard de la deuxième thèse ci-dessus. Dans le second, ses explications sont insuffisantes.

Ces remarques ne permettent cependant de répondre qu'à une partie de l'objection wittgensteinienne, et montrent qu'il est parfaitement cohérent de supposer qu'il existe des processus de traitement d'information à un niveau subintentionnel, qui sont pertinents pour l'explication des processus intentionnels. Mais elles ne permettent pas de répondre à l'autre partie de l'objection wittgensteinienne: pourquoi ces processus subintentionnels ou subpersonnels de traitement d'information seraient-ils des processus traitant une information

⁴ Stone et Davies (*ibid.*, p. 614) font une pétition de principe contre les wittgensteiniens quand ils disent que les deux explications sont en compétition.

sémantique , c'est-à-dire des processus véhiculant certains contenus? Pourquoi ne seraient-ce pas des processus purement neurophysiologiques, qui n'ont à ce titre aucun caractère *cognitif* ? Si c'était le cas, quelle raison aurait-on de parler de *neuropsychologie* et qui plus est de neuropsychologie *cognitive* ? Si l'on veut respecter la position de Shallice, il faut dire que la neuropsychologie cognitive, comme la psychologie cognitive, postulent l'existence de traitements d'information sémantique, et donc en ce sens mentale et intentionnelle, mais inconsciente. Or c'est précisément l'idée qu'il puisse exister une telle information qui est mise en cause par les wittgensteiniens. Ce dont ils doutent est l'idée qu'il puisse exister un ordre de représentations mentales intermédiaires entre ce que Marr appelle le niveau computationnel et le niveau physique. Selon eux, ou bien ces représentations se réduisent à des états neurophysiologiques, auquel cas elles n'ont plus rien de "mental", ou bien elles n'existent tout simplement pas (pour des arguments semblables, voir Searle, 1992).

Là encore, cet argument s'appuie sur l'idée selon laquelle un état mental représentationnel doit avoir, nécessairement, un certain nombre de caractéristiques; c'est une affaire de tout ou rien: ou bien ces caractéristiques sont exemplifiées, et l'état est "mental" ou elles ne le sont pas, et c'est une pure mythologie que de supposer qu'il pourrait exister des états mentaux qui n'auraient pas ces caractéristiques. Mais ce raisonnement repose sur une pétition de principe. En l'occurrence la pétition de principe est qu'il existe un certain nombre de conditions nécessaires pour qu'un état mental soit un état mental. Mais qui décide ici de ces conditions nécessaires? Apparemment c'est le philosophe. Mais quelle garantie avons-nous que sa description est correcte? Je voudrais essayer de montrer qu'il n'existe aucune garantie de ce genre quand nous confrontons le concept philosophique et de sens commun d'*état mental conscient* avec divers cas bien connus où une certaine forme de conscience est absente, mais où une autre forme de conscience est présente. Ces cas ne peuvent pas être simplement décrits comme des cas où la conscience est absente, ni par conséquent comme des cas où aucun

état mental est présent. Cela conduit à réviser la notion ordinaire de conscience sur laquelle s'appuient les philosophes et le sens commun.

Les cas en question sont bien connus. Ce sont ceux de vision aveugle et de commissurotomie (Shallice, 1995: 484-510). Ce qui m'intéresse est la description et l'interprétation qu'en donnent les philosophes. Commençons par la vision aveugle. Des patients atteints d'une lésion dans leur cortex visuel primaire ont typiquement des aires "aveugles" dans leur champ visuel. Les patients rapportent ne rien voir d'un stimulus visuel (par exemple un flash de lumière) dans leur aire atteinte, mais sont capables de deviner correctement certains traits du stimulus, comme par exemple des formes. Clairement la conscience, sous une forme quelconque, manque à ces patients. Le problème est de savoir quelle forme de conscience manque, et quelle forme de conscience est présente. Quelle forme de conscience? Apparemment la forme de conscience qui permet de déployer des informations dans un raisonnement et contrôler rationnellement l'action. Par exemple, Marcel (1986) observe qu'un patient assoiffé victime de vision aveugle n'était pas capable d'atteindre un verre d'eau dans son champ visuel aveugle. Il est tentant de soutenir ici que puisque la conscience est absente dans la vision aveugle, la conscience doit avoir comme fonction de permettre à des informations représentés quelque part dans le cerveau d'être utilisées dans des raisonnements et de guider rationnellement l'action. L'idée est que quand un contenu n'est pas conscient il peut influencer le comportement de diverses manières, mais que seul un contenu conscient peut jouer un rôle rationnel, et que seule la conscience est capable de promouvoir ce rôle rationnel. Searle (1995: 155) propose un argument du même genre à partir des observations de Penfield au sujet d'épileptiques qui sont atteints d'une crise de petit mal au moment où ils effectuent des actions telles que marcher, jouer du piano ou conduire et continuent à effectuer ces activités en dépit du fait qu'ils ne sont pas conscients. Il en conclut qu'une fonction de la conscience est de promouvoir la flexibilité et la créativité du comportement. Le cas de la vision aveugle ou du petit mal a ceci de commun avec les cas de prosopagnosie que les

patients atteints de prosopagnosie sont aussi capables de “deviner” certaines propriétés des visages qui leur sont présentés, même s’ils ne sont pas capables de les catégoriser correctement. Le fait que la conscience soit la partie de l’esprit qui facilite le contrôle rationnel de l’action n’est pas incompatible avec certains modèles généraux du fonctionnement cognitif, comme celui de Schacter “DICE” (Dissociable Interactions and Conscious Experience, 1989).

- insérer ici figure 1 -

Selon ce modèle, la conscience existe en fait à deux niveaux: comme conscience “phénoménale” et comme conscience “exécutive”. La première sert de point de passage pour la seconde, qui contrôle directement le raisonnement et l’action. Une fonction de la conscience, selon ce modèle, inclut l’intégration des sorties des modules spécialisés et la transmission des contenus en question aux mécanismes de contrôle de l’action. Selon une interprétation plausible de ce modèle, les cas de prosopagnosie et de vision aveugle peuvent se comprendre comme des cas où le système d’entrée de reconnaissance des visages ou le système visuel ont été déconnectés du système exécutif. Cela reste parfaitement compatible, notamment avec la conception fodorienne de la modularité selon laquelle les systèmes sensoriels périphériques sont isolés par rapport au système central de production de croyances. Mais il y a une autre interprétation possible des données. Le raisonnement de Searle, selon lequel les patients sont victimes de vision aveugle suppose que la conscience est absente dans ces cas. Mais quelle conscience? Il est parfaitement plausible de supposer qu’il existe ici deux sortes de conscience : une conscience phénoménale, qui n’est pas associée à un contrôle de l’action mais seulement à la perception de qualités phénoménales et une conscience associée à un contrôle de l’action. C’est en fait l’hypothèse que propose le philosophe Ned Block (1995), qui distingue ce qu’il appelle conscience phénoménale (“P-conscience”) et conscience comme accès (“A-conscience”). La

première serait relative à la perception des expériences, des sensations, alors que la seconde serait relative à l'accès, réflexif, que nous pouvons avoir aux propriétés représentationnelles de ces états. Selon Block l'une des consciences peut exister sans l'autre, et c'est donc une erreur que de supposer que si l'une des consciences manque dans les cas de vision aveugle, alors l'autre manque aussi. Les deux concepts distincts de conscience sont ici confondus, selon Block. Un exemple clair de ce genre de confusion est commis, selon lui, par Searle dans des passages comme le suivant, où il discute des cas de patients épileptiques décrits par Penfield:

“Dans tous ces cas, nous avons affaire à des formes de comportement finalisé sans la moindre conscience. Or pourquoi les comportements ne pourraient-ils être tous ainsi? Qu'est-ce que la conscience ajoute?... La conscience ajoute des pouvoirs de discrimination et de flexibilité même aux activités de routine mémorisées”.

(Searle, 1995: 155)

Searle soutient officiellement que la conscience est la conscience phénoménale, et qu'il y en a des degrés. Mais ici il est clair qu'il utilise la notion de conscience en un autre sens, celui de la conscience comme accès. Il ne s'ensuit pas que l'on puisse dire nettement que, dans les cas de vision aveugle en particulier, on a affaire à un type de conscience plutôt qu'un autre. Le raisonnement de Searle est un cas typique où on part d'une prémisse *a priori* selon laquelle on a affaire à un type d'état mental, alors qu'il pourrait en exister au moins deux types. Searle se repose ici, comme de nombreux philosophes, sur ce qu'Anthony Marcel appelle “l'hypothèse de l'identité”, selon laquelle “les représentations qui constituent l'expérience consciente sont les mêmes exactement que celles qui sont dérivées et utilisées dans le traitement de l'information sensorielle (Marcel, 1983: 238). L'hypothèse de l'identité est naturelle si l'on s'en tient à la psychologie du sens commun. Quand nous rapportons un contenu d'attitude propositionnelle que nous avons ou que quelqu'un d'autre a, nous présupposons que le *même* contenu se trouve dans la pensée de l'individu et dans celui du rapporteur. C'est même la raison pour laquelle nous ne pouvons pas nous tromper quand nous avons une pensée au sujet de nos

pensées: si je dis “Je crois que p ”, je peux me tromper sur le contenu p (par exemple je peux être victime d’une illusion visuelle), mais je ne peux pas me tromper sur le fait que j’ai une pensée dont le contenu est p . Cette transparence cartésienne, comme on pourrait l’appeler, n’est pas niable. Mais il n’y a aucune raison de supposer qu’elle fait partie de tous les états mentaux.

Le second exemple que je voudrais donner provient d’une recherche de Justine Sergent (1990). Sergent traite de patients ayant subi une commissurotomie et auxquels on transmet différentes informations à chaque moitié de cerveau. Ici on projette des paires de chiffres simultanément, un chiffre à chaque hémisphère; et on demande aux patients de les comparer en poussant un levier avec l’une ou l’autre de leurs mains. Dans la première expérience, la tâche consiste à comparer les deux nombres et à dire s’ils sont le même ou différents. Quelle que soit la main qu’ils utilisent, les patients fournissent des réponses qui ne sont pas meilleures que si elles étaient données au hasard. Mais dans une seconde expérience, on demande aux patients si un nombre est plus élevé (plus grand) que l’autre. De manière surprenante, les patients trouvent cette tâche “très aisée” et répondent correctement sans hésitation. C’est étonnant, car si 7, par exemple, est plus élevé que 6, cela implique qu’ils ne sont pas le même nombre, et donc la réponse à la première question. Dans une troisième expérience, les patients sont capables d’indiquer, correctement et précisément, si la paire de nombres qu’ils comparent sont égaux; mais ils sont incapables de dire, quand la première expérience est reprise une heure plus tard, si les nombres sont les mêmes ou différents. L’interprétation naturelle de ces résultats est qu’une information au sujet d’une comparaison de quantités est d’une façon ou d’une autre partagée entre les deux hémisphères, mais pas l’information portant sur l’identité des nombres, ni sur la forme ou le nom de ces nombres. L’absence d’unité affecte certaines catégories d’informations, mais pas d’autres. Ici aussi on peut se demander, comme dans le cas de la vision aveugle: ont-ils une conscience ou non de ces nombres?

La philosophe Susan Hurley (1993) propose une interprétation “ kantienne ” de ces résultats. Kant peut être compris comme défendant l’idée, dans l’*Analytique transcendantale*, selon laquelle l’unité de la conscience requiert l’objectivité, c’est-à-dire quelque chose d’extérieur à la conscience⁵. Il existe un certain sens de la notion d’unité de la conscience que l’on peut formuler ainsi : si une pensée que p et une pensée que q ont lieu dans la même conscience, alors il est naturel d’inférer que, simultanément avec la pensée que p et la pensée que q , il y a la pensée que p et q , ce que l’on peut appeler *principe de transitivité* (ou d’agglomération). Il semble essentiel aux pensées conscientes, quand elles sont ainsi conjointes, qu’elles puissent être scindées en pensées distinctes, appartenant chacune à son “ monde ”, c’est-à-dire qu’on puisse leur assigner des indices pour les distinguer. En particulier si la pensée p et q est contradictoire, nous devons assigner des indices différents aux pensées correspondantes (par exemple si je pense que vous êtes malade, et que vous n’êtes pas malade, ces deux pensées doivent ne pas être simultanées, ou avoir un indice distinct). Dans le cas des sujets atteints de commissurotomie, il semble que nous ne puissions donner un sens à leur comportement que si nous faisons l’hypothèse que le principe précédent est violé, c’est-à-dire que s’ils pensent que p , et pensent que q , ils ne pensent pas que p et q . Si un patient de Sergent a un ‘6’ projeté sur son hémisphère gauche, et un ‘7’ sur son droit, et qu’aucun côté n’est capable de rapporter de manière fiable si les nombres sont les mêmes ou différents, implique que la perception de ‘6’ n’est pas co-consciente avec la perception de ‘7’. Mais le fait que les deux côtés sont capables de rapporter que le nombre sur la gauche est plus petit que celui sur la droite peut être pris comme impliquant que la conscience du nombre sur la gauche comme plus petit est co-consciente avec la perception de ‘6’, et aussi avec la conscience du nombre sur la gauche comme plus grand, qui est co-conscient avec la perception de ‘7’. La co-conscience n’est pas une relation

⁵ C’est l’interprétation, par exemple de P.F. Strawson, dans son livre *The Bounds of Sense* (Methuen, London, 1966). Peu importe ici que cette interprétation soit correcte ou non, puisque ce qui nous intéresse c’est le passage en général d’un argument purement philosophique à un argument fondé sur la neuropsychologie de la conscience.

transitive. La conclusion de Hurley est qu'il n'est pas possible de distinguer l'unité de la conscience, dans ces cas, sur la base de ressources seulement subjectives. Il doit y avoir une unité objective, extérieure au sujet, et extérieure au contenus de consciences eux-mêmes. Mais cette unité n'est que faible ; elle ne peut pas être aussi forte que le principe de transitivité.

Ici on a une conclusion transcendantale – l'unité de la conscience ne peut pas être une unité purement subjective, elle doit dépendre du monde objectif dont on est conscient – qui paraît tirée d'observations neuropsychologiques. Cela paraît être bien plus que ce que la neuropsychologie n'est habituellement capable de montrer. Il semble en fait que tout ce que montrent les données de Sargent, si elles sont correctement interprétées, est qu'il y a des cas où notre idée intuitive de l'unité de la conscience est battue en brèche. Mais on a le sentiment que son argument prouve trop.

Quelles conclusions méthodologiques tirer de ces quelques exemples de relations entre philosophie et neuropsychologie ? J'ai distingué au début deux manières de considérer l'une comme pertinente pour l'autre : soit que les données neuropsychologiques viennent à l'appui de certaines conclusions philosophiques, soit que certaines élucidations conceptuelles philosophiques permettent de clarifier des problèmes théoriques en neuropsychologie. Un sceptique peut contester cette pertinence dans les deux directions. Il peut soutenir que les analyses conceptuelles et a priori quant à, par exemple, la notion de conscience, ne seront jamais touchées par les moindres données neuropsychologiques. Des argumentations comme celle que je viens d'évoquer au sujet des cas de commissurotomie pourraient nous conduire à aller dans ce sens. Mais ce scepticisme n'est pas justifié quand on considère les analyses de la connaissance tacite que j'ai examinées en premier lieu. Notre sceptique peut aussi soutenir que l'on ne peut jamais tirer la moindre vérité significative philosophique de données neuropsychologiques. Ici encore ce scepticisme est battu en brèche. La conclusion n'est pas qu'il faut prêcher en la

matière un œcuménisme de bon aloi, mais que même s'il n'est pas évident que des disciplines comme la philosophie et la neuropsychologie puissent être pertinentes l'une pour l'autre, tout verdict négatif en la matière est prématuré.

Références

- BLOCK N. (1995), On a confusion about the function of consciousness, *Behavioural and Brain Sciences*, 18, 227-287.
- Coltheart, M. (1985), Cognitive neuropsychology and the study of reading, In M.I. Posner & O.S.M. Marin (eds), *Attention and Performance*, Hillsdale, N.J., L. Erlbaum, 11,
- DE HAAN E.H.F., YOUNG A.W., NEWCOMBE A. (1987), Face Recognition without awareness, *Cognitive Neuropsychology*, 4, 385-415.
- DESCOMBES V. (1995), *La denrée mentale*, Paris, Minuit.
- ENGEL P. (1996) *Philosophie et psychologie*, Paris, Gallimard.
- ETCHOFF N.L., FREEMAN R., CAVE K. (1992), Can we lose memory of faces? Content specificity and awareness in prosopagnosia, *Journal of Cognitive Neuroscience*, 3,
- FODOR J. (1983), *The modularity of mind*, Cambridge, Mass., MIT Press, trad. Fr. *La modularité de l'esprit*, Paris, Minuit, 1987.
- HURLEY S. (1993), Unity and objectivity, in C. Peacocke (ed.), *Objectivity, simulation, and the unity of consciousness*, British Academy & Oxford University Press.
- MALCOLM N. (1959), *Dreaming*, Londres, Routledge.
- MARCEL A. (1983), Conscious and unconscious perception: An approach to the relations between phenomenal experience and perceptual processes, *Cognitive Psychology*, 15, 238-300.
- SCHACTER D.L. (1989), On the relation between memory and consciousness : dissociable interactions and conscious experience, In H.L. Roediger & F.I.M. Craik (eds), *Varieties of memory and consciousness*, Hillsdale, NJ, L.E.A.
- SEARLE J. (1992), *La redécouverte de l'esprit*, Paris, Gallimard, 1995.
- SELLARS W. (1963), *Science, perspective and reality*, London, Routledge.
- SERGENT J. (1990), Furtive incursions into bicameral minds, *Brain*, 113, 538-568.
- SHALLICE T. (1995), *Symptômes et modèles en neuropsychologie*, Paris, PUF.
- STONE T., DAVIES M. (1994), Cognitive neuropsychology and the philosophy of mind, *British Journal for the Philosophy of Science*, 44, 589-522.
- YOUNG A.W., DE HAAN E.H.F. (1988), Boundaries of overt recognition in prosopagnosia, *Cognitive Neuropsychology*, 5, 317-336.

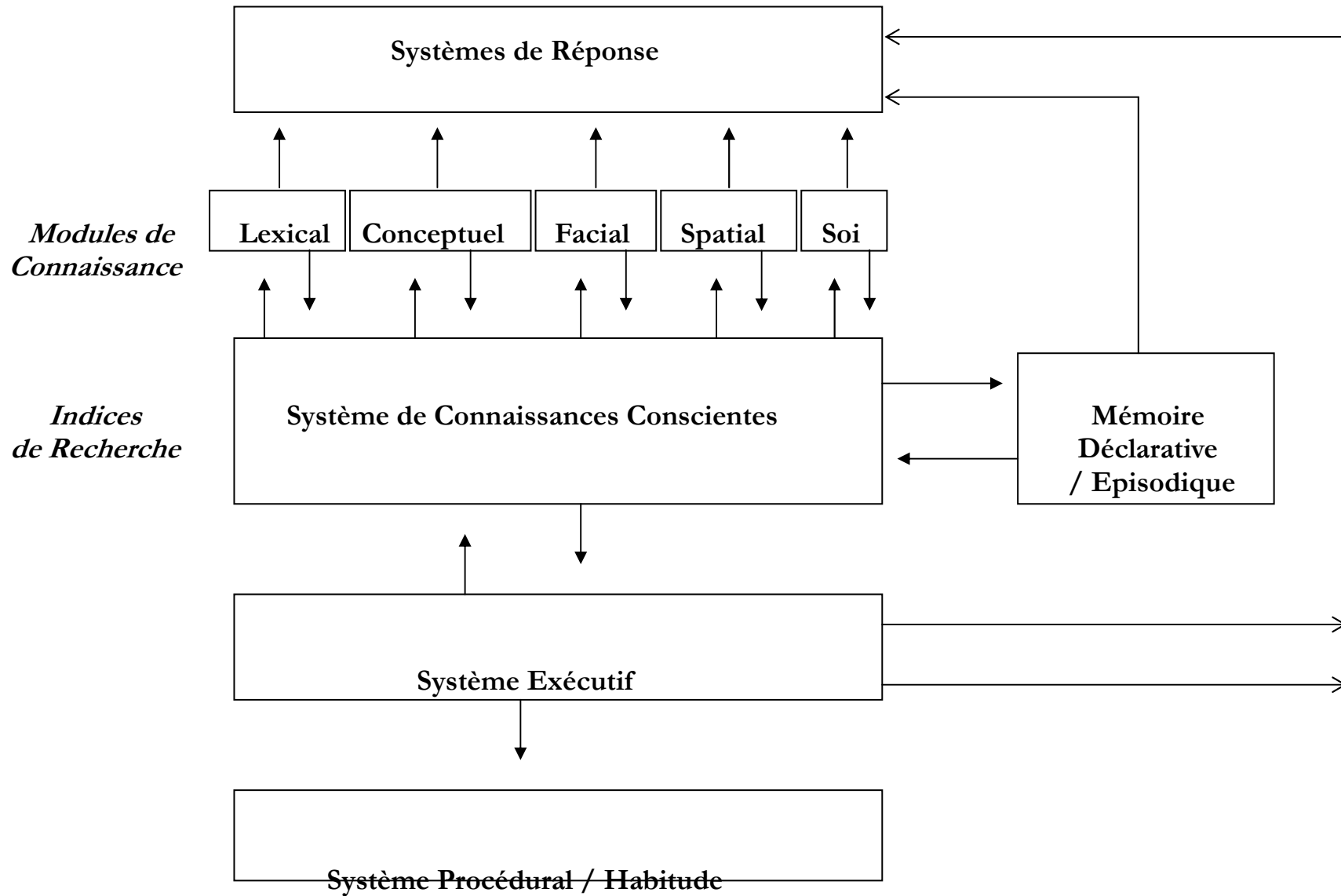


Figure 1 : Modèle DICE d'après Schacter (1989).